

Use of Mixed Stated and Revealed Preference Data for Crowding Valuation on Public Transport in Santiago, Chile

Marco Batarce, Juan Carlos Muñoz, Juan de Dios Ortúzar, Sebastián Raveau, Carlos Mojica, and Ramiro Alberto Ríos

The valuation of comfort on public transport is presented with mixed stated preference and revealed preference data. In this case, comfort is measured mainly as the level of crowding in the vehicles (bus or train) with the use of in-vehicle passenger density (in number of passengers per square meter). The data used to value comfort include a stated preference survey in which crowding levels are presented as illustrations and revealed preference data on route choice on the subway network of Santiago, Chile. The survey data are used to estimate discrete choice models and obtain a subjective valuation of passenger density through the parameters of the utility function. Disutility for traveling in crowding conditions is assumed to be proportional to the travel time; therefore, the longer the trip, the higher the utility loss. Results indicate that passenger density has a significant effect on the utility of public transportation modes. In fact, marginal disutility of travel time in a crowded vehicle (6 passengers/m²) is twice the marginal disutility in a vehicle with a low level of crowding (1 passenger/m²).

In the transport planning literature, one of the best solutions—or perhaps the best one—to the widely recognized problem of traffic congestion and air pollution derived from urban transport is high-quality mass public transit. Many cities already have implemented high-capacity transit systems (bus rapid transit, tramway, or subway). However, the engineering and operational designs of these systems often does not consider adequately the effect of comfort on travel demand. This oversight leads to a design standard of ≥ 6 passengers/m², which is exceeded at peak times in some corridors. In addition, crowding is omitted from most demand models used for planning public transport. Therefore, a more detailed understanding is needed of crowding in mass transit systems and its impact on travel decisions.

The mode choice process is complex because it considers many more variables than those usually included in traditional demand models: fare, in-vehicle travel time, and waiting time (1). Empirical evidence indicates that the comfort levels of different alternatives may be significant factors in explaining travel behavior (2–4). Obviously

associated with comfort, crowding also can imply a perceived lack of control or stimulus overload, among other stressors. Thus, travel decisions involving comfort and crowding are complex mental processes that involve attitudes, psychological states, preferences, and socioeconomic constraints.

For that reason, the question of how to measure crowding is relevant for transport policy and planning. The general objective of this study is to analyze crowding and comfort attributes with mixed stated preference (SP) and revealed preference (RP) data. Conducted in Santiago, Chile, the study focuses on estimating comfort valuation measured as travel time.

Almost all works that address the valuation of comfort in public transport systems use SP methods; most use choice-based SP methods (5). However, some SP studies also have applied contingent valuation to find the willingness to pay for reducing overcrowding (6, 7). In a review of public transport crowding valuation research, Li and Hensher focus on studies conducted in the United Kingdom, the United States, Australia, and Israel (5). Most of these studies use logit models with choice-based SP data covering commuters and focus mainly on in-vehicle congestion costs. Douglas and Karpouzis estimate crowding costs at the platform (related to waiting time) and at the access or entrance to train stations (related to walking time) (8).

In choice-based SP surveys, the crowding representation is relevant. Whelan and Crockett use the seat occupancy rate and the number of standing passengers as a proxy for crowding to conduct an SP experiment (9). These parameters allowed calculation of the load factor (number of passengers per number of seats) and passenger density (standing passengers/m²) and specified time multipliers according to each level. It means that the effect of crowding in the modal utility is represented as a factor that increases the negative perception of travel time. Wardman and Whelan suggest that passenger density is a better indicator of congestion, given that a same load factor may have different levels of crowding across different types of trains with different seating configurations (10).

In addition to the time multiplier approach, crowding cost can be obtained directly as a monetary value. Lu et al. conducted an SP experiment for train users in 2005 and estimated values for crowding costs that were more than twice as high as the value of in-vehicle time in an uncongested scenario (11). In the survey, crowding was represented as the combination of a probability of occurrence and length of time (e.g., two out of five times someone stands for the whole 30-min journey).

The present study uses disaggregated mixed SP and RP data. In choice-based SP surveys, one attribute of the alternatives is related to crowding. The crowding level is supported by pictures that depict

M. Batarce, Industrial Engineering School, Universidad Diego Portales, Vergara 432, Santiago, Chile. J. C. Muñoz, J. de Dios Ortúzar, and S. Raveau, Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile. C. Mojica and R. A. Ríos, Department of Infrastructure and Environment, Inter-American Development Bank, 1300 New York Avenue, NW, Washington, DC 20577. Corresponding author: S. Raveau, sraveau@puc.cl.

Transportation Research Record: Journal of the Transportation Research Board, No. 2535, Transportation Research Board, Washington, D.C., 2015, pp. 73–78.
DOI: 10.3141/2535-08

vehicle passenger densities. The RP data are observed route choices in the subway network of Santiago, Chile, collected in a survey. Therefore, the mixed data comprise not only observed and hypothetical choices but also transport mode and route choices.

In the city of Santiago, 11.5 million motorized trips occur daily, of which about 40% are on public transport. The numbers of people who use the bus system and the subway are similar; each mode transports about 2.3 million passengers each day. The average in-vehicle travel time for a trip on public transport is 28.5 min, and the average travel distance is 12 km; average transit speeds are 17 km/h for bus trips and 33.4 km/h for subway trips.

The rest of the paper is structured as follows. The data used for estimation are presented in the next section. Then, the modeling approach based on discrete choice modeling is described, and the results of estimation are presented. To conclude, some final comments and policy implications are presented.

DATA

SP data were collected in an experiment that considers six choice scenarios with two alternatives each. Six attributes for each alternative are described: transport mode, travel time, travel cost, average waiting time, waiting time variability (i.e., coefficient of variation), and crowding level in the vehicle (bus or train).

Transport modes were presented to individuals according to the real availability in the reported trip. For a respondent who used public transportation, the choice scenarios included only bus and train as alternatives. For a respondent who used a car, he or she was asked whether public transportation was a travel option. In this case, if the respondent can use only bus, the choice scenario includes car and bus; otherwise, the choice scenarios include car versus bus or train. If the respondent cannot use public transportation, the choice scenarios are completely hypothetical, including only public transportation alternatives. Depending on mode availability, four optimal designs were built with Ngene software from ChoiceMetrics (12). All choice scenarios included the response, "I would not travel in any alternative," as recommended in the literature (1). Batarce et al. describe the SP survey in detail (13).

The level of attributes is determined according to the actual level reported by a respondent in a reference trip. Levels of travel time were pivoted on the actual travel time of the longest leg of the reference trip. Travel cost level depended on the transport mode. For bus or train modes, travel cost levels were 590 Chilean pesos (CLP) (US\$1.00) or 650 CLP (US\$1.20). When the alternative was car, travel cost levels were 100% or 110% of the actual travel cost computed on the basis of travel time, average speed, and average fuel consumption. If the respondent paid for parking or an urban highway toll in the reference trip, these expenses were added to the car travel cost but only in the final survey design.

Average and variability of waiting time levels were different for every mode. Average waiting time was 0 min for car, 5 or 10 min for bus, and 3 or 5 min for train. Waiting time variability is measured with the coefficient of variation, which was 0 for car and 0, 0.5, 0.7, 1.0, and 1.5 for public transportation. This attribute was presented in the survey as a range of time during which the next bus or train arrives with uniform probability.

Six crowding levels for bus and train were presented to survey respondents as illustrations (Table 1). Each illustration was associated with a level of crowding from 1 to 6 passengers/m². The survey was developed on a computer and conducted in person by a surveyor.

The survey took 15 to 20 min to complete and therefore was carried out at the workplaces of the respondents. Figure 1 is a choice scenario presented to the respondents.

The survey form also included questions about the respondent (i.e., gender, age, car ownership, income) and a reference trip to work. The reference trip included travel time, frequency, and legs of the trip and then, for each leg, the mode, in-vehicle travel time, waiting time, comfort level (sitting, standing with room around, standing with little room, or standing in an overcrowded vehicle), and parking and toll costs, if any. This information is used to pivot the attributes presented in the choice scenarios.

RP data come from real (i.e., observed) choices made by travelers on the Santiago Metro underground railway network (4). The alternatives are the potential routes to be taken between the travelers' origins and destinations, of which several variables differ (e.g., in-vehicle travel time, waiting time, walking time when transferring, number of transfers, and crowding level on trains). These data were obtained from an origin–destination survey conducted at Metro stations in October 2008 that gathered information about the trips of 92,800 individuals (about 12% of all users). Only those whose origin–destination pairs could have been taken by more than one route are included in the study analysis, leaving 28,961 individual choices to be used for estimating the models. These data were complemented by detailed information about the system levels of service, with values obtained for all explanatory variables. Crowding levels for the RP database were obtained from observed load profiles in the different Metro lines. In this RP database, the fare is not included as an explanatory variable because monetary cost is the same within the Metro network regardless of route length or transfers (and thus has no effect on route choice).

MODELING AND ESTIMATION

Model Specification

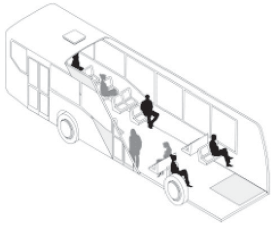
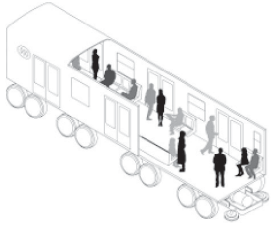








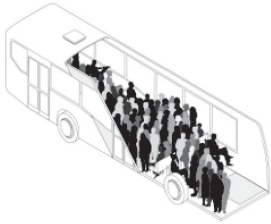
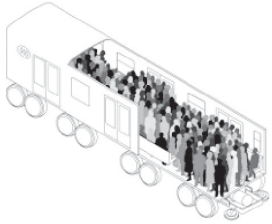
The framework for model specification is the random utility theory. In the context of transport mode choice, the random utility theory can be summarized in the following assumptions about an individual's behavior:

- A (finite) set of mutually exclusive transportation alternatives is available for the trip.
- Individual preferences about alternatives can be represented by a utility function that depends on the attributes of the alternatives and characteristics of the individual.
 - The individual chooses the mode that generates the highest utility among all available alternatives in the choice set.
 - In the utility function, some variables can be observed only by the individual. This way, two individuals with the same choice set and the same observable characteristics may choose different transportation modes.
 - The unobservable individual utility components are assumed to be random and independently distributed in the population.

The random utility component comes from different sources (e.g., any unobservable or unmeasurable attribute of the alternatives or unobserved individuals' taste variation).

In practical terms, the theory of random utility involves defining a utility function for each mode, which has as variables modal attributes, individual characteristics, and a random component that distributes

TABLE 1 Passenger Density and Figures Used to Represent Level of Crowding

Level of Crowding (passengers/m ²)	Bus	Train
1		
2		
3		
4		
5		
6		

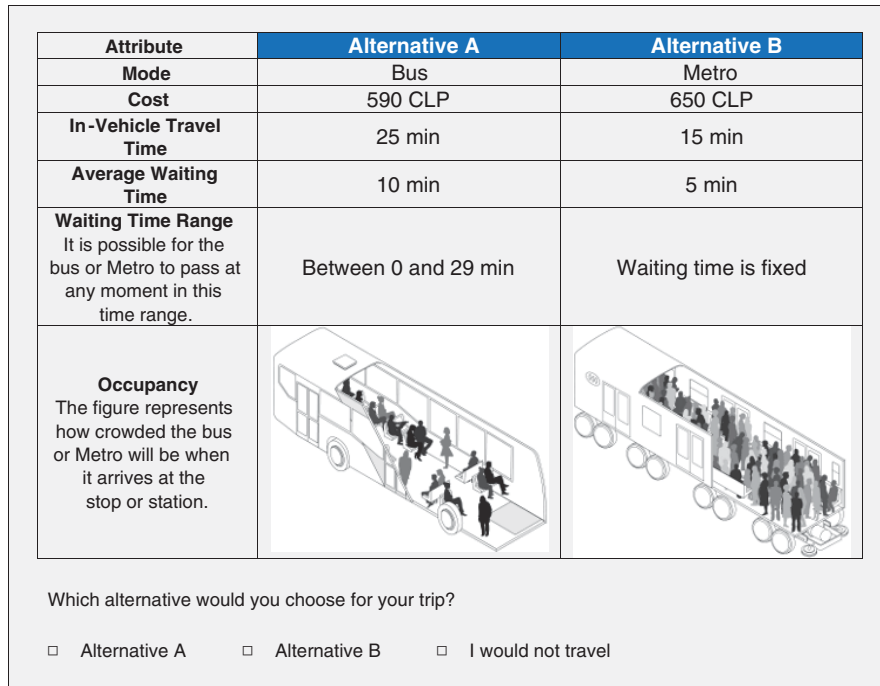


FIGURE 1 Proposed SP choice scenario.

over the population. Analytically, the random utility of alternative m for individual i is written as $V(x_m, z_i, e_{mi})$, where x_m is a vector with m mode attributes (travel time, cost, and so forth), z_i is a vector with characteristics of individual i (income, driver license, and so forth), and e_{mi} is the random component. This utility also is a random variable.

Because individuals are assumed to choose the alternative that maximizes their utility, the mode m is chosen if $V(x_m, z_i, e_{mi}) \geq V(x_k, z_i, e_{ki})$ for all k modes in the set of modes available to the individual. Because the utility is a random variable, the probability that individual i chooses alternative m is $\text{prob}(i \text{ chooses } m) = \text{prob}(V(x_m, z_i, e_{mi}) \geq V(x_k, z_i, e_{ki}) \text{ for all } k)$. Different models are obtained depending on the assumptions made about the functional form of the utility and the probability distribution of the random component. The logit model assumes that the additive random component is separable in the utility function and Gumbel distribution.

Moreover, the logit models generally assume that the observable part of the utility function is linear in mode attributes. Therefore, if the utility function includes time and cost, then the associated parameters represent the marginal utility of such variables. For example, if $V_i = a_i + bC_i + cT_i + e_i$, where a_i is an alternative specific constant, C_i is travel cost, and T_i is travel time, then b is marginal utility of income and c is marginal utility of time (in simplified terms).

Because the marginal rate of substitution between money and time corresponds to the subjective value of time, this value can be calculated as the ratio of the time and cost parameters of the linear utility function. This study assumes that the marginal utility of travel time depends on the crowding level in the vehicle; this approach is consistent with the time multipliers approach (5). Therefore, the marginal utility of travel time is multiplied by passenger density (7). This specification captures the increasing discomfort of traveling under crowded conditions and implies that total discomfort is proportional to travel time.

In summary, the discrete choice models are logit models with linear utility functions and time multipliers to capture crowding effects. The choice probability is given as the following:

$$P_{im} = \frac{e^{\lambda V_{im}}}{\sum_k e^{\lambda V_{ik}}} \tag{1}$$

and the utility functions are given by the following:

SP data:

$$V_m = a_m + bC_m + \sum_j c_j D_{jm} T_m + dWT_m \tag{2}$$

RP data:

$$V_m = a_m + \sum_j c_j D_{jm} T_m + dWT_m + fW_m + gTR_m \tag{3}$$

where

- P_{im} = probability of individual i choosing alternative m ,
- λ = scale parameter related to random component variance,
- C_m = cost of mode m ,
- D_{jm} = dummy variable for passenger density j ,
- T_m = travel time,
- d = marginal utility of waiting time,
- WT_m = waiting time,
- W_m = walking time,
- f = marginal utility of walking time,
- g = marginal utility of transferring, and
- TR_m = transfers in the subway network.

The crowding effect is modeled with dummy variables to capture a nonlinear effect. Also, the RP utility function does not include the cost of the alternative because data correspond to individuals making route choices within the subway network without having to pay extra for transferring between lines. Therefore, all alternatives have the same cost. Socioeconomic variables (especially the respondent’s gender) interact with the explanatory variables but are not significant. One important assumption is that the car crowding level is equal to the lowest crowding level on public transport (i.e., 1 passenger/m²).

The λ parameter in Equation 1 is the scale factor and measures the variance of the error term in the utility function. This factor usually is not identifiable with a unique sample of individuals and therefore is normalized to 1. However, when estimation data come from different samples, the scale factors can be identified for every sample except one. The difference between samples depends on the nature of the data. For instance, data coming from SP and RP or data from SP surveys with different experimental designs have different natures and, therefore, choice probabilities with different scale factors (14).

The SP survey comprises four experimental designs, differentiated by specific scale factors. The RP survey also has a specific scale factor to differentiate it from the SP survey.

Estimation Results

The estimation sample includes 3,380 choice situations corresponding to 580 individuals who were surveyed with four final experimental designs. These design differences imply that the logit model estimation should consider different scale factors. For identification, one scale factor is normalized to 1. The RP database consists of 28,961 choices.

Table 2 summarizes estimation results for three proposed models: SP data, RP data, and joint SP and RP data. Even though the parameter associated with waiting time uncertainty was presented in the SP survey and its effect in the utility function was estimated, it is not shown because this work focuses on the direct effect of crowding on time valuation. To prevent the bias on other parameters of the utility

because of wrongly specified waiting time uncertainty, the effect of this parameter was modeled with dummy variables.

All explanatory variables are significant on the three estimated models with different data. Only the Metro constant is not statistically different from the bus constant. Marginal disutility of in-vehicle travel time increases as the level of crowding increases. Waiting and walking times present a higher disutility than in-vehicle time (because of uncertainty and physical effort, respectively).

The utility specification considers a nonlinear effect of crowding on travel time with dummy variables to represent the levels. Preliminary estimations considered specific parameters for each level of crowding; however, some are statistically equal each other. Therefore, the crowding effect is aggregated into three levels that are associated with three travel conditions. In the low-density level (1 to 2 passengers/m²), a transit user has a clear opportunity to sit. In the medium-density level (3 to 4 passengers/m²), users travel standing but with some space. In the high-density level (5 to 6 passengers/m²), users travel standing and overcrowded. The nonlinearity in the utility function indicates the great discomfort of traveling under overcrowded conditions in contrast to traveling seated or standing with few passengers in the vehicle.

Results indicate that crowding increases travel disutility significantly. Marginal disutility increases 29% when passenger density increases from 1 to 2 passengers/m² to 3 to 4 passengers/m² and increases 73% when passenger density increases from 3 to 4 passengers/m² to 5 to 6 passengers/m². One minute of traveling under the high-density condition (5 to 6 passengers/m²) produces discomfort 2.3 times greater than that produced in the low-density condition (1 to 2 passengers/m²). For policy design, decreasing crowding from the third level to the second may affect demand and user welfare significantly.

Table 3 presents the marginal rates of substitution between variables (notably, values of time) for the full-data, joint SP and RP model to further analyze individual perceptions. The value of in-vehicle travel time varies from 2,626 CLP (US\$4.60) to 5,894 CLP (US\$10.40) per hour depending on levels of crowding. Valuations for waiting time and walking time are higher. Individuals are willing to pay 250 CLP (US\$0.44) to avoid a transfer.

TABLE 2 Model Estimation for Santiago with Mixed Data

Parameter	SP		RP		SP and RP	
	Estimate	t-Test	Estimate	t-Test	Estimate	t-Test
Monetary cost	-0.001	-3.67	—	—	-0.0008	-4.37
Travel time at 1–2 passengers/m ²	-0.042	-6.41	-0.117	-51.17	-0.035	-9.24
Travel time at 3–4 passengers/m ²	-0.054	-8.41	-0.132	-56.65	-0.045	-8.87
Travel time at 5–6 passengers/m ²	-0.091	-11.71	-0.194	-43.99	-0.078	-8.63
Waiting time	-0.098	-9.69	-0.183	-8.24	-0.079	-13.18
Walking time	—	—	-0.257	-13.00	-0.076	-7.65
Transfers	—	—	-0.698	-10.26	-0.241	-5.13
Bus constant	0.000	FV	—	—	0.000	FV
Metro constant	0.017	0.21	—	—	0.031	0.44
Car constant	1.64	2.39	—	—	1.93	8.37
SP scale factor, Design 1	1.000	FV	—	—	1.000	FV
SP scale factor, Design 2	0.692	3.62	—	—	0.692	3.62
SP scale factor, Design 3	1.150	9.35	—	—	1.150	9.35
SP scale factor, Design 4	0.519	4.05	—	—	0.519	4.05
RP scale factor	—	—	1.000	FV	3.821	8.84

NOTE: — = variable was not part of model specification; FV = variable was fixed to ensure model identification. Sample size: SP = 3,380; RP = 28,961; SP and RP = 32,341; log likelihood: SP = -1,870; RP = -13,480; SP and RP = -15,609; corrected ρ²: SP = .567; RP = .382; SP and RP = .403.

TABLE 3 Marginal Rates of Substitution for Santiago with Joint SP and RP Model

Parameter	Valuation
Travel time at 1–2 passengers/m ²	2,626 CLP/h
Travel time at 3–4 passengers/m ²	3,389 CLP/h
Travel time at 5–6 passengers/m ²	5,894 CLP/h
Waiting time	4,903 CLP/h
Walking time	4,642 CLP/h
Transfer	250 CLP/transfer

Other travel conditions also are significant in the modal utility; waiting time is valued as high as the travel time in overcrowded conditions. The same is true for walking time in transfers. This result implies that comfort improvements also may focus on walking time reductions in transfers. In addition, the need to transfer to reach a destination produces significant disutility. In fact, without considering the required walking and waiting time, one transfer is valued as 6 min of in-vehicle travel time—that is, 263 CLP (US\$0.46).

FINAL COMMENTS

This paper used mixed SP and RP data to value the effect of crowding on public transport. SP data were collected in a survey based on mode choice that included public transport and car; RP data were collected in a survey of subway users who chose routes on the Metro network. In the SP survey, the level of crowding was measured as in-vehicle passenger density and presented to respondents as images. In the RP survey, routes differ in travel time, number of transfers, crowding levels, walking time in transfer, and waiting time. Therefore, users are assumed to consider these trip variables in making their choices.

Discrete choice models were used to value crowding and specify modal utility function in which passenger density increases the effect of travel time on the utility. In other words, it supposed an interaction between passenger density and travel time. The results of parameter estimation show that crowding has a significant and nonlinear effect on the marginal utility of travel time. Indeed, marginal disutility of travel time in a vehicle with 6 passengers/m² is twice that of a vehicle with 1 passenger/m².

Methodologically, the approach in this paper is innovative in mixing data on modal choice with data on route choice. This joint SP and RP model not only benefits from the advantages of both preference approaches but also allows the valuation of more attributes related to comfort (which were not shown in the SP experiment to prevent a cognitive burden when respondents completed the survey).

One policy implication of this study is that transit passengers perceive the effect of crowding similarly to how motorists perceive road congestion. Improving the travel time of a bus line increases demand for that service. Consequently, this new demand increases crowding, and in turn, the generalized travel cost (travel disutility). These two effects balance each other, which may present challenges to effectively implementing transport policies that are intended to increase the operation speed of public transport without increasing capacity to avoid crowding. Two examples of this effect are the bus rapid transit system in Bogotá, Colombia, and the Metro system in Santiago. Both cities reduced the generalized travel costs in travel time or travel fare (Bogotá reduced travel time by implementing TransMilenio, and Santiago reduced the Metro fare by integrating

the fare scheme with the bus system). As a result, the in-vehicle passenger density in both systems is higher than the design density. This high crowding might prevent car users from being attracted to public transportation.

Results of this study can be used to include cost of crowding (or congestion) on public transport during the planning and appraisal stages of public transport projects. Cost–benefit analyses could benefit from better estimation of the technical and economic feasibility of transit investments. Moreover, planners and policy makers might be able to prepare adequate finance plans for increased vehicle capacity provisions intended to control the negative effects of crowding.

Finally, if users consider the level of crowding when choosing a line or route of public transport, the final demand of each line is the result of an equilibrium state. This equilibrium is similar to that in a road network with traffic congestion, which implies the need to develop transit network assignment models that consider an effect similar to road congestion on bus routes. This topic should be included in future research.

REFERENCES

- Ortúzar, J. de D., and L. G. Willumsen. *Modelling Transport*. 4th ed. John Wiley & Sons, Chichester, United Kingdom, 2011.
- Ben-Akiva, M., J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou. Integration of Choice and Latent Variable Models. In *Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*. Elsevier, Oxford, UK, 2002, pp. 431–470.
- Cherchi, E., and J. de D. Ortúzar. Mixed RP/SP Models Incorporating Interaction Effects. *Transportation*, Vol. 29, No. 4, 2002, pp. 371–395.
- Raveau, S., J. C. Muñoz, and L. de Grange. A Topological Route Choice Model for Metro. *Transportation Research Part A: Policy and Practice*, Vol. 45, No. 2, 2011, pp. 138–147.
- Li, Z., and D. A. Hensher. Crowding and Public Transport: A Review of Willingness to Pay Evidence and Its Relevance in Project Appraisal. *Transport Policy*, Vol. 18, No. 6, 2011, pp. 880–887.
- Guerra, G., and J. Bocarejo. Congestion Cost in Mass Transit Systems; Pricing and Investment Policy Implications—Case Study: Bogotá's BRT System. Presented at 13th World Conference on Transportation Research, Rio de Janeiro, Brazil, July 2013.
- Haywood, L., and M. Koning. *Discussion Papers 1293: Estimating Crowding Costs in Public Transport*. Deutsches Institut für Wirtschaftsforschung, Berlin, 2013.
- Douglas, N., and G. Karpouzis. Estimating the Passenger Cost of Station Crowding. *Proc., 28th Australasian Transport Research Forum (ATRF)*, Sydney, New South Wales, Australia, 2005.
- Whelan, G., and J. Crockett. An Investigation of the Willingness to Pay to Reduce Rail Overcrowding. *Proc., 1st International Conference on Choice Modelling*, Harrogate, UK, 2009.
- Wardman, M., and G. Whelan. Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. *Transport Reviews*, Vol. 31, No. 3, 2011, pp. 379–398.
- Lu, H., T. Fowkes, and M. Wardman. Amending the Incentive for Strategic Bias in Stated Preference Studies: Case Study in Users' Valuation of Rolling Stock. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2049, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 128–135.
- Rose, J. M., and M. C. J. Bliemer. Constructing Efficient Stated Choice Experimental Designs. *Transport Reviews*, Vol. 29, No. 5, 2009, pp. 587–617.
- Batarce, M., J. C. Muñoz, and J. de D. Ortúzar. Valuing Crowding in Public Transport Systems: The Case of Santiago. Presented at 13th Pan-American Conference on Transportation, Traffic and Logistics, Santander, Spain, June 2014.
- Ben-Akiva, M., and T. Morikawa. Estimation of Travel Demand Models from Multiple Data Sources. Presented at 11th International Symposium on Transportation and Traffic Theory, Yokohama, Japan, 1990.